Eco. Env. & Cons. 30 (4) : 2024; pp. (1888-1898)
Copyright@ EM International
ISSN 0971–765X

DOI No.: http://doi.org/10.53550/EEC.2024.v30i04.070

Genome-wide *In-silico* Identification of Microsatellites in Eri Silkworm, *Samia ricini* (Lepidoptera: Saturniidae)

Vanitha C.[1&2], Indumathi Kamatchi Balakrishnan[1&2], Rajal Debnath[1], Lavanya C.[1&2], Tulsi Naik K.S.[1], S. Manthira Moorthy[1], Kureeckal V. Ramesh[2] and Himanshu Dubey[1*]

[1]Seri-Biotech Research Laboratory, Central Silk Board, Govt. of India, Kodathi, Bangalore 560 035, India
[2]Department of Biotechnology, School of Sciences, Jain University, Bangalore 560 011, India

(Received 3 July, 2024; Accepted 13 August, 2024)

ABSTRACT

Effective utilization of microsatellite or simple sequence repeat (SSR) markers along with ongoing advanced technologies in the field of molecular biology have made great advantages. In the present study, the genome-wide distribution of SSRs in Eri silkworm, *Samia ricini* Donovan was studied along with the *in-silico* transferability of SSR primers to the closely related species *S. wangi* and *S. watsoni.* We also analyzed and compared the SSRs in two closely related *Samia* species with the motifs identified in *S. ricini* genome. The overall abundance of SSR motifs was more in *S. watsoni* compared to *S. ricini* and *S. wangi.* The cross-amplification analysis through *in-silico* PCR method with the 25,252 SSR primer pairs developed for *S. ricini*, showed that 2,978 and 1,508 primer sets were amplified in the *S. wangi* and in *S. watsoni*, respectively. The distribution of SSR motifs in *S. ricini* genome showed a higher proportion of SSR motifs in intergenic regions than exonic and intronic regions (41%, 30% and 29%, respectively). The gene ontology analysis of the SSR containing genes showed that a greater number of genes were associated with the biological process (889) followed by molecular function (215) and cellular components (213). The SSRs present in the coding region revealed, Alanine (Ala) as the most abundant amino acid in these loci followed by Tyrosine (Tyr) and Methionine (Met) in the coding region of the *Samia ricini*. To validate our *in-silico* analysis, we randomly selected 20 SSR primers and amplified them in different morphotypes of Eri silkworm.

Key words: *Microsatellite, Simple sequence repeats, Eri silkworm,* Samia ricini

Introduction

Eri silkworm, *Samia ricini* Donovan (Lepidoptera: Saturniidae), is a holometabolous lepidopteran insect (Fig. 1) that belongs to the family Saturniidae. Eri silkworm is a semi-domesticated Vanya silkworm also known as ahimsa silk, endi or errandi in India. It is raised in India and parts of the orient for its silk (Ravinder *et al.,* 2016). In India, Eri silkworm holds a major share in Vanya silk production of 7364 MT over other vanya silks *viz.,* tasar and muga silkworms, 1466 MT and 255 MT, respectively (https://texmin.nic.in/documents/annual-report). In total 2.94 lakh families are involved in the ericulture in India having 33,433ha of land covered under ericulture (Kumara, 2023). The other serigenus insect species belong to the genus *Samia viz., S. wangi* and *S. watsoni* and are not exploited for commercial

(*Scientist-C)

silk production. Though *S. watsoni* is believed to be a sister taxon to all other *Samia,* a lot of ambivalence was aroused previously, whether to continue its genera in the same or to move into a new genus. The phylogenetic studies imply that *S. ricini* is having closer relationship with *S. wangi* than the *S. watsoni* (Lu *et al.,* 2022a, b). The geographical distribution of *S. ricini* can be seen in East Asian whereas *S. wangi* in Southeast China, Taiwan and Northern Vietnam and *S. watsoni* in South-eastern, Asia and China (https://www.silkmothsandmore.com). *S. ricini* has been considered as a new model species other than *Bombyx mori* in molecular and cellular research (Hosamani *et al*., 2015; Maegawa *et al*., 2018) with the advantages of tolerance and easier rearing in addition to the silk yield compared to *Bombyx mori* (Chakrabarty *et al*., 2012).

A genetic marker is a variation in allele that can be used for rapid identification of breeding lines, hybrids, cultivars and species, facilitate genetic diversity and relatedness estimations in germplasm and allows phylogenetic relationships to be established with more accuracy than was previous possibilities of morphological and biochemical techniques (Wani, *et al.,* 2013). Microsatellites or Simple sequence repeats are the genomic sequences highly utilized in the population genetics as a molecular marker. These are comprised of tandem repeats of short nucleotide motifs (Mets *et al.,* 2016 and Ranade *et al.,* 2015) containing repeats of a motif sequence, 1-6 bp in length and influenced by the base composition of the flanking sequence (Prasad *et al.,* 2005). The polymorphic SSR loci provides excellent targets and means of assessing genetic variation in samples (Foster *et al.,* 2011). Over the years, advances in molecular genetics methodology have led to widespread use of co dominant molecular markers, especially SSRs. In addition, they are utilized in breeding programs for marker-assisted selection, which quickens the process by cutting down most of time. Microsatellites are extensively employed in genetics studies, using both low and high throughput genotyping approaches (Vieira *et al.,* 2016).

The role of SSRs in different studies like analysis of genetic diversity (Mason A. S., 2005 and Mishra, A. *et al.,* 2022) through genetic distance in Tasar silkworm, *Antheraea mylitta* (Renuka and Shamitha, 2016), gene flow estimation (Pérez-Martínez *et al.,* 2022), rate of crossing over and in evolutionary studies , also to infer intraspecific genetic relations between the closely evolved species (Mandal, 2018) and in addition to these the SSRs also useful for the construction of linkage map in *Bombyx mori* (Zhan *et al.,* 2009), mapping of loci involved with QTLs in *Bombyx mori* (Li *et al.,* 2013), the degree of kinship between genotypes (Madesis *et al.,* 2013), in marker assisted selection for characterizing the transposable elements in the in *Bombyx mori* (Osanai-Futahashi *et al.,* 2008) and defining cultivar DNA fingerprints (Kalia *et al.,* 2011).

The *Bombyx mori* microsatellites feature average distances of between 49 and 161 kb (Vieira *et al.,* 2016) with genome coverage is 0.31 and 16% and they have a cloning efficiency of between 0.77 and 3.5% (Miao *et al.,* 2005). Together, such values indicate that microsatellites in the *B. mori* genome are rare. The pattern is not unusual however, having been observed in other insect species, including several in the Drosophila genus (Meglécz *et al.,* 2007; Van't Hof *et al.,* 2007). Being a model organism for insect studies, *B. mori* silkworm has been studied expansively by researchers for more than two decades using different molecular markers. Especially in genetic fingerprinting of silkworm genotypes and as a mapping tool, ISSR-PCR method is potentially useful since strain-specific pattern was shown to be inherited and segregated in a Mendelian fashion in silkworms (Reddy *et al.,* 2000). The linkage mapping of silkworm genes like of *nlw*, a recessive gene was studied based on the SSR linkage map and STS markers (Wang *et al.,* 2010) showed that molecular markers are of greater use and undeniable for practical usage. The phylogenetic relationships of different ecoraces attained on the basis of RFLP pattern support the phenotypic and geographical relations in tropical Tasar silkworm, *Antheraea mylitta* Dury (Botlagunta *et al.,* 2006) and the genetic diversity of different Tasar ecoraces were carried out by using SSR markers (Renuka and Shamitha, 2016). Likewise, ISSRs markers were used to study the genetic diversity and differentiation among populations of the Indian Eri silkworm *S. ricini* (Vijayan *et al.,* 2006), Miao *et al.,* (2008), used the fluorescent SSR markers to construct a linkage map for the female heterogametic silkworm (*B.mori*, ZW). Also looking into the tremendous advantages of the SSR molecular markers in an organism, there is a wider scope of these markers in the Eri silkworm improvement. Thus, the obtained data from this study can be utilized for the improvement of the Eri silkworm breeds not only in terms of silk yield also to develop a breeds or hybrids for more pupal weight because of its greater

feed value for poultry and fisheries even for human consumption in some parts of India (Assam).

Unlike other silkworms, the Eri silkworm exhibits a noteworthy tolerance for diseases. It shows reduced susceptibility to diseases unless the pathogen load is significant, suggesting a robust resistance mechanism (Lu *et al.,* 2022a), the details of which are still to be fully revealed. The findings of the current study, particularly the SSR markers, offer a promising solution to this issue. They hold the potential for identifying and validating the genes associated with disease resistance in the Eri silkworm in future research endeavors. The current study yields valuable insights into SSR markers, which can be effectively utilized in breeding programs. These markers offer crucial information that can contribute to the enhancement and optimization of breeding initiatives.

Studying SSRs in any genome of an organism gives the genetic variability across the species level. Information on SSRs are well documented in other sericigenous insects (Wang *et al.,* 2010; Behura and Severson, 2012; Renuka and Shamitha, 2016) but lack in Eri silkworm, though the knowledge on SSRs are well known to dates back to 1990's (Jurka and Pethiyagoda, 1995). Recently Eri silkwormis emerging as a new model species in lepidopteran insects (Lee *et al.,* 2021), shows the need for the current work. Thus, keeping Eri silkworm, *S. ricini* as a main impetus, we studied the abundance of SSRs along with the two other species and examined the cross-amplification ability of obtained primers from *S. ricini.* In our study the SSR markers were designed for the *S. ricini*, to test their effectiveness in amplifying the target sequences of *S. wangi* and *S. watsoni* since they are related or similar species (Lu *et al.,* 2022a). This is the first of its kind study reporting the SSR motif abundance and distribution in the Eri silkworm, *Samia ricini.*

Materials and Methods

Genome wide dentification of SSR markers

The whole-genome sequence data for the Eri silkworm, *S. ricini* was retrieved from the NCBI database (Gen Bank assembly accession**:** GCA_014132275.1) of size 450.47Mb of length. To identify SSRs in *S. wangi* and *S. watsoni*, raw genomic reads were downloaded from the SRA database of NCBI deposited under the PRJNA818466 - SRR18441011 and PRJNA818861 - SRR18441626 respectively. Quality analysis and removal of adapter and low quality bases of downloaded reads were carryout using FastQC (Andrews, 2010) and Trimmomatic tools (Bolger *et al.,* 2014). The SPAdes version 3.15.5 was used for the genome assembly *S. wangi* and *S. watsoni* using filtered high-quality reads. The sequences were then analyzed to identify the perfect SSR loci using the MIcroSAtellite (MISA) software (Beier *et al.,* 2017) following the set to detect tandem repeats of di-, tri-, tetra-, penta- and hexa-nucleotides, with a minimum number of 6, 5, 5, 5 and 5 tandem arrays of the core repeat, respectively. Further the identified SSRs were classified as class I (≤20 bp) and class II SSR (12-20 bp). The compound SSR where perfect SSRs are interrupted by non-repetitive nucleotides were removed for the further analysis. The contingence ÷2 test was used for testing heterogeneity of SSR motif counts in all the three species of *Samia.* The abundance of the different SSR motifs and the AT and GC % of SSR motifs were subjected for the chi square test. The statistical analysis was performed in R software.

SSR primer designing

The SSR primer sets were generated from Primer 3.0 (Untergasser *et al.,* 2012) software for the *S. ricini* genome. The minimum SSR motif length of ≤ 18 nucleotideswere considered for the SSR primer pairs designing. The other parameters like melting temperature (5G5Z) of 55 - 64 °C with optimum annealing temperature of 58 °C, optimum GC content 40% and amplicon length of 120- 1000 bp were modified from a default configuration.

Cross amplification study through *In-silico* PCR

Considering the advantage of SSR markers, cross amplification studies were conducted through *in-silico* PCR method to check the transferability primers generated for *S. ricini* to the *S. wangi* and *S. watsoni* genomes. A pearl scriptin_silico_PCR.pl obtained from (https://github.com/egonozer/in_silico_pcr/blob/master/in_silico_PCR.pl) was used for this analysis without allowing any mismatches between the primer sequence and the target genome

Gene ontology and amino acid distribution

To analyze the distribution of SSRs in the genic region of *S. ricini* we first performed the gene annotation for this genome using GenSAS webserver (Humann *et al.,* 2019) using transcriptome files ob-

tained from NCBI-SRA (SRA accession no. DRR213133, DRR213134, DRR213141, DRR213142, DRR213143, DRR213144) database as hint. The overlap between genes and SSR loci was analyzed through BED tools (Quinlan and Hall, 2010). Functional annotation of the genes overlapping with SSR loci, was performed through diamond blast (Buchfink, 2021) and InterproScan (Jones *et al.*,. 2014) and the result was combined using BLAST2GO (Conesa *et al.*, 2005). The genes which are linked to SSRs loci were analyzed for functional categorization through REVIGO (Supek *et al.*, 2011).

Isolation of genomic DNA

Genomic DNA was isolated as described in other studies (Ismail *et al.*, 2020) with minor modifications from the six different morphotypes (A: Yellow Plain, B: Yellow Spotted, C: Yellow Zebra, D: Greenish Blue Plain, E: Greenish Blue Spotted, F: Greenish Blue Zebra) of Borduar eco race of Eri silkworm, *S. ricini.* The individual moth crushed with micro pestle, immediately added 500µl of 2PK buffer (200mM Tris Ph-8.0, 25mM EDTA, 300mM NaCl, 2% SDS) in 1.5 ml vial to completely homogenize the tissue. The homogenized sample then subjected for the centrifugation at 12000 rpm for 10 min at room temperature to get a clear supernatant devoid of cell debris. The supernatant was transferred to a fresh vial and equal volume of phenol chloroform isoamyl alcohol mixture (25:24:1) was added followed by centrifugation. The supernatant was then transferred to fresh tube and added equal amount of chloroform isoamyl alcohol (24:1) mixture. The aqueous layer was collected into a fresh 1.5ml vial and equal volume of chilled isopropanol was added and the tube was inverted several times to precipitate the genomic DNA and centrifuged at 12000 rpm for 10 min at 15°C to pellet the genomic DNA followed by DNA wash with 500 µl of 70% ethanol, then pellet was air dried. The DNA pellet was resuspended in 50µl of autoclaved Mili-Q water and the quality of DNA was analyzed on 0.8% agarose gel.

Validation of SSR markers

The SSR primers designed for *Samia ricini* were synthesized from Eurofins Genomics India Pvt. Ltd., Bengaluru. The PCR amplification performed using 10 µl of PCR mixture solution containing 5 ml of 2X Emerald (TaKaRaBio,), 1 µl of 10 pmols each primer 50 ng of genomic DNA template. PCR amplification was run. with the following program: 94 °C for 5 min, followed by 35 cycles of 94°C for 40 s, then 40 s at the annealing temperature of each primer pair, 72 °C for 1 min 30 s, and a final extension at 72 °C for 10 min in T1000™ Thermal Cycler (Bio-Rad Laboratories, Inc.). PCR products were separated on 1.5% agarose gel using a Maxi Submarine Gel System (Bioworld Pvt. Ltd., Bengaluru), stained with ethidium bromide and photographed under UV light using Omega Flour™ Gel documentation System.

Results and Discussion

SSR frequency distribution in three *Samia* species

The abundance of SSR motifs was analyzed in all three species of genus *Samia* and divided into class I and class II SSR based on their motif length. The overall number of the SSR motifs was observed more in the *S. watsoni* followed by *S. ricini* and *S. wangi* (Table 1). In Class I, dimers are most abundant in *S. ricini* and *S. wangi* whereastetramer and hexamers are the least abundant in all three species of Samia (0.22, 0.23 and 0.20 %, respectively). In Class II, dimers are most present in *S. ricini, S. wangi* and *S. watsoni* (43.58, 51.76 & 39.85%, respectively) compare to trimers which are least in number (24.72,

Table 1. The SSR distribution in *S. ricini, S. wangi* and *S. watsoni.*

	Samia ricini	*Samia wangi*	*Samia watsoni*
Genome size (Mb)	450.47	329.03*	831.80*
No. of contigs (Million)	-	0.3	2.6
%GC	34.3	36.72	39.03
Total length of the SSRs (bp)	768350	346860	1983733
Number of SSRs	35622	20890	98725
Relative abundance of SSRs	79.08	63.49	118.69
Relative density of SSRs	1705.66	1054.18	2384.86

*In house assembled data through SPAdes version 3.15.5

23.96 & 20.92%, respectively) (Table 2). In *S. ricini* the order of abundance in class I SSRs was dimer> tetramer> trimers> pentamer> hexamers while in class II it was dimer> trimers. The abundance of the class II SSRs were same in all three species. Likewise, in *S. wangi* the order was dimers> tetramers> trimers> pentamers> hexamers in class I SSRs, while it was tetramer> dimer> pentamers> trimers> hexamers in *S. watsoni* (Table 2 and Fig. 2a). The overall SSR motif number as well the length of SSR motifs were found to be more in the *S. watsoni* followed by *S. ricini* and *S. wangi* (Fig 2a). The total number and length of the SSR motifs in the current study are less when compared with the earlier studies in *Bombyx mori* may be because of the strict tandem array of the core repeat foreach motif (Reddy *et al.*, 1999) for each motif. In addition, the class I SSRs are more in number in comparation with the class II is also because of this strict assign of the tandem array for the SSR motifs. Out of which the dimer motifs (abundance %) found be more in all species (65.30%, 66.22% and 53.42 % in *S. ricini, S. wangi* and *S. watsoni,* respectively), followed by trimer> tetramer> pentamer> hexamer motifs, which is in parallel with the earlier reports in *Bombyx mori* (Reddy *et al.,* 1999). There was a significant difference among the frequencies for each motif when compared among the three species of the *Samia*, showing a greater deviation from the expected frequencies except for the hexamers in *S. wangi,* which was found to be non-significance (chi-square = 7, p-value = 1). The greater number of combinations were found in pentamers followed by tetramers> hexamer> trimer> dimer (Table 2). These variable repeats are distributed differently in different genome sequences indicated that the nucleotide composition, the distinctive DNA replication, repair, and recombination machinery may have had a significant influence in the development of SSRs among all the species of genus *Samia.* The TA dimer motif found to be most frequent in all the three species for both class I and class II SSRs which is similar to the earlier reports in the *Bombyx mori* (Prasad *et al.*, 2005), where TAA was found to be more abundant among the trimers only in class I of *S. ricini.* In *S. wangi* and *S. watsoni* having ATA is the most frequently found trimer motif whereas earlier reports in *Bombyx mori* showed GCA, CGA and CGG were most represented than other trimer motifs (Prasad *et al.*, 2005). In tetramer, the TTTA is the most frequently occurred in *S. ricini* and *S. wangi,* whereas AGAC found to be most frequent in *S. watsoni.* AACCT is the most abundant pentamer motif in *S. ricini* and *S. wangi* where GGTTA is more frequent pentamer in *S. watsoni* (Table S1, Fig 2b). Therefore, these variations in frequency of SSR motifs are either analysis is dependent on data sets, or the frequency might be specific to the species and the parameters employed in this study for SSR mining might be having the effect.

The top 200 utmost SSRs in each species for both Class I and Class II are enumerated in the Table S1 (Fig. 2b). The differential counts for SSR in class I and class II revealed that, the AT% was found to be more when compared with GC% in all the species of *Samia* and there was no significant difference between the three species.*S. watsoni* had recorded lower AT% than *S. wangi* and *S. ricini* in class I SSR,

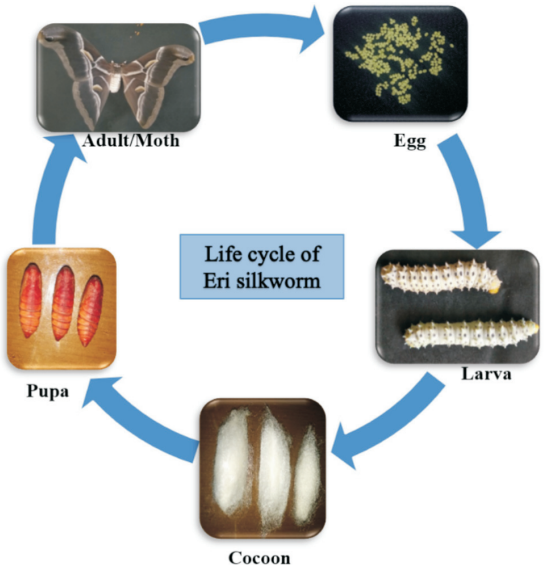

Fig. 1. Life cycle of holometabolas Eri silkworm *Samia ricini.*

Table 3. GC and AT % across the three species of Genus *Samia.*

	S. ricini		*S. wangi*		*S. watsoni*	
	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2
GC%	7	15	16	15	24	12
AT%	89	77	80	77	72	80

Table 4. List of the primer sets selected for the validation of SSR primers in *Samia ricini*.

Sl. No.	SSR IDs	SSR	Primer binding Chromosome	Forward primer	Reverse	Tm primer	Amplicon size
1.	SR00499	(AT)89	BLXV01000001.1	AGCGGCTGTCTTTTTGTTTGTT	ATTGTGACCAAAGCACTAAGCC	57	374
2.	SR06922	(AT)83	BLXV01000003.1	CCAACTCAGACTAGAATTTTCA TGCA	AGACTTATGCAGATACAGCACCT	59	290
3.	SR08486	(AT)94	BLXV01000004.1	TAATACCAGCGAATACTGCCCG	GCGTACTCACATCCCGAAATTG	60	398
4.	SR13695	(TA)88	BLXV01000006.1	ACTTGGTAATTATATCCGGGCGT	TGACGTACGTATCAATTATGTGTGA	58	357
5.	SR16385	(TA)94	BLXV01000007.1	AGAGTGTTTGATGACCGGTACC	GAATAGCAAAAACGGCGAGGAG	60	423
6.	SR18034	(TA)89	BLXV01000008.1	TTGCAATGGCCACCAATAGAAC	TTGTTATCCTTAAGCTGGCACC	58	281
7.	SR18319	(AT)86	BLXV01000009.1	TGTTGATTTGATCTCTGGGAGGA	AGCCTGCTCAGAATACTATATGAACT	59	310
8.	SR22918	(TA)76	BLXV01000012.1	GTTGTCACTTGTGTAGCAGCTG	ATGTTTACGACGGTCAGTGTGA	59	225
9.	SR24779	(AT)76	BLXV01000013.1	AACAACGTGTTCATCGCCAAAT	GTTTGTCTGGACGAGAAGGGAA	58	218
10.	SR31824	(CA)59	BLXV01000023.1	CACAATCAGATGCCACAGAACC	AACACGAGGAAAGCTTTTGACG	59	281
11.	SR32749	(TA)82	BLXV01000025.1	CGGGATAACCTTCGCTTTTCAC	ACTGAGCATCGTGTACGTGTTA	59	432
12.	SR29630	(TA)78	BLXV01000018.1	CGGACTCCGTTCGCTTATTCTA	CTTAACATCAGGTGCAGTGCAG	60	310
13.	SR21055	(AT)77	BLXV01000010.1	ACCTCGAGTTGTTTAGTCGCAA	TCGGCCGTTAAATATTTCAAACAAC	58	278
14.	SR11203	(AT)76	BLXV01000005.1	GAGGTTATGGTGCCTCGAAGAA	GGTGTCTTTTGAATGGCAGCTT	59	269
15.	SR13852	(TA)76	BLXV01000006.1	AACAGACAGACATGCAGACTGA	TGTCAGCGGATTGTCATTTTGA	57	276
16.	SR30853	(TA)76	BLXV01000020.1	CGCAGGCACAGCTAGTTTAAAA	CGATTTACGAACACACACTGTTTTGA	58	370
17.	SR00375	(AT)75	BLXV01000001.1	TGGGTACTCGACACATTAAACCGT	TTTGTCATTAAAACAGACGCTTCAA	57	330
18.	SR08430	(GCG)5	BLXV01000004.1	TTCAGAGGGTTGAGATGATGGC	GGTGAGAGGAGGAGGATGTACA	61	253
19.	SR25625	(CAG)5	BLXV01000014.1	GCAACCCTACCCCGAGTATTAC	GCGAGTGTGCCACAAAGTATAG	61	271
20.	SR33037	(CAA)5	BLXV01000026.1	TCATTCACCGACCAGAAACGAT	TGGCTGAACAAAATGCTGGTTT	57	227

however in class II, the former had higher AT% than the latter two. The GC% observed to be more in *S. watsoni* than *S. wangi* and *S. ricini* in class I SSR whereas in class II it was more in *S. wangi* and *S. ricini* than the *S. watsoni* (Table 3). These findings are on par with the earlier results where the AT motifs were more abundant compared to GC in the *Bombyx mori*. The earlier reports it was evident that the AT motifs were found to be rich than the GC motifs and the % of AT and GC were positively related to the flanking sequences (Prasad *et al.*, 2005).

The genome wide distribution of, 35,622 SSRs *S. ricini* showed 20,501 SSR fall under the intergenic region and 15,117 are under genic region. In the genic region 455 microsatellites overlapped with exonic and UTR regions and 14,662 are in intronic regions (Fig. 2c). The variation of SSR distribution in the whole genome of *S. ricini* was studied at different regions of the genome. In general, the SSR abundance at different regions varies greatly may be because of variation in the replication slippage occurs when the DNA polymerase slips or stutters during replication, leading to the misalignment of the template and the newly synthesized strands resulted in the formation of the SSRs/ microsatellites. The SSRs located at the CDS region of the genome greatly affect the gene expression of an organism, not only the exonic SSRs but also the intronic SSRs are believed to be involved in the gene transcription, gene splicing, mRNA splicing. Not only that the SSRs present at the 52UTRs and 32UTRs are also involved in the regulation of gene expression and can cause the transcriptional slippage (Qi *et al.*, 2016). The distribution of these SSRs at CDS regions seems limited by non-perturbation of the reading frame. SSRs play an important role in organism devel-

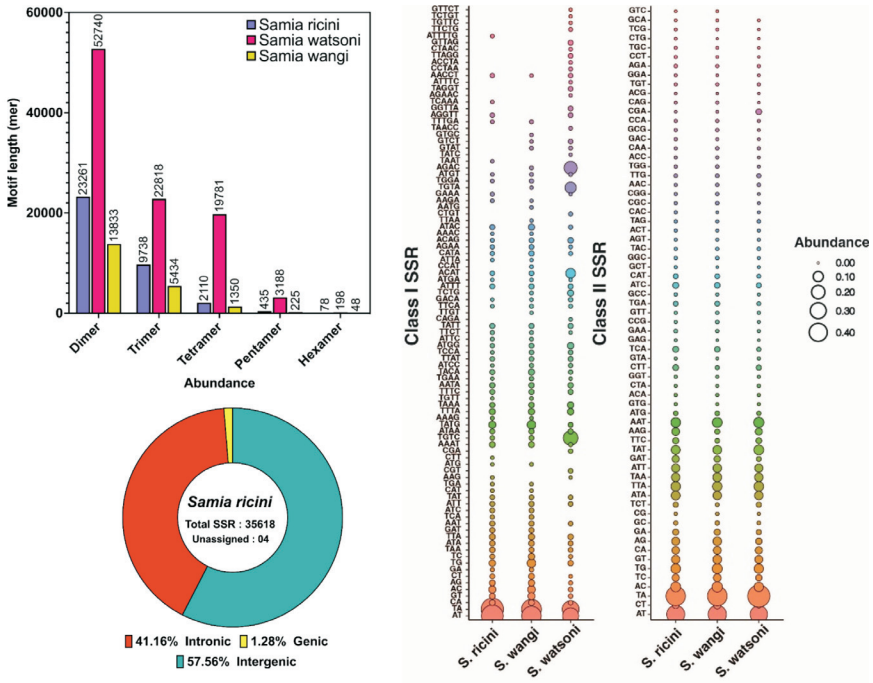

Fig. 2. The SSR distribution. (a) The perfect SSR motifs in *Samia ricini, S. Watsoni* and *S. wangi.* (b) The top 200 SSR motifs across three genomes of *Samia* in both class 1 and class 2. (c) Distribution of the perfect SSR motifs in *Samia ricini* genome.

opment, survival, adaption and so on (Li *et al.,* 2004). The (CAG)n SSR motif in the Homeobox gene (DLX6) could lead to cell death in the apical ectodermal ridge causing limb malformations in the Drosophila spp. (Ferro *et al.,* 2001). In present study a total of 35622 SSRs were found *S. ricini* and most of them were distributed at the intergenic (20,501) followed by genic region (15,117) In genic SSRs 455 are observed to be at exonic and UTR regions and 14,662 are present at the intronic regions (Fig 2c). The similar observations were reported by the earlier studies (Behura and Severson, 2012).

Cross amplification among the different species of Genus *Samia*

The transferability of SSR primers generated for *S. ricini* was tested against the *S. wangi* and in *S. watsoni* genomethrough *in-silico* PCR. Over all 25,252 primer sets with a single amplicon (Table S2) were selected for the *in-silico* PCR through the pearl script. Among the amplified sets it is observed that 235 primer sets are common to both the species whereas 2753 in *S. wangi* and 1273 in *S. watsoni* wereunique to respective species. The species considered in this study are belongs to the same genus *Samia* have believed to be evolved from the common ancestry. This proves that, they are closely related species which is supported by the earlier studies (Lu *et al.,* 2022b).

Functional annotation of SSR containing Genes and Amino acids encoded by microsatellite at coding region

The gene prediction was conducted for the whole genome of *S. ricini*, which was unavailable in the NCBI database and the gene prediction statistics of *Samia ricini* is given in Table S3. Afterward, a homology search against the NCBI-nr database was performed using Diamond Blast, and additional domain information on these genes were extracted utilizing InterproScan. As a result, Eri silkworm has got 18,078 genes and 13,873 genes are functionally annotated, whereas only 455 genes found to be having SSR motifs out of which 363 are functionally annotated (S5 and S6). The gene ontology (GO) analysis was executed by REVIGO tool to study the functional analysis of these genes comprehending the SSRs in *Samia ricini*. The results showed that,the greater number of genes were associated with the biological process (889) followed by mechanical function (215) and cellular components (213) in *S. ricini*. In the present study more number of genes

having SSRs were involved in the biological process than the cellular components and molecular function which is supported by earlier reports (Hill *et al.*, 2016; Aamir *et al.*, 2017; Kim *et al.*, 2017). The cellular process (GO:0009987), metabolic process (GO:0008152), primary metabolic (GO:0044238), nitrogen compound metabolic process (GO:0006807) and macromolecule metabolic process (GO:0043170) were found to be abundant in biological process. However, cellular anatomical entity (GO:0110165), membrane components (GO:0016020) and intracellular anatomical structural components (GO:0005622) were showed over expression in cellular components whereas ion binding (GO:0005488; GO:0043167; GO:0043168; GO:0046872), transferase activity (GO:0016740), nucleic acid binding (GO:0003676), hydrolase activity (GO:0016787) and nucleotide binding (GO:0000166) are found to be abundant in the molecular function (Fig 3 and Table S4). For better understanding of the functional annotation, the scatterplots were drawn which are resultant of GO terms obtained after the redundancy reduction. The scatterplot indicates cluster representatives in two-dimensional space obtained by applying multidimensional scaling to a matrix of the GO terms' semantic similarities which explains larger the bubbles more the general terms (Fig. 3).

Alanine (Ala) was the amplest amino acid in in both class I and class II SSR motifs in *Samia ricini.* Proline (Pro) was the second most occurred amino acid followed by Histidine (His)in class I SSR motifs while Arginine (Arg) most abundant next to Alanine (Ala) followed by Glycine (Gly) in case of class II SSR motifs. Tyrosine (Tyr) is the least abundant amino acid in class I and Methionine (Met) in the class II are the least occurred one (Fig4).The alanine is most frequently occurred amino acid encoded by microsatellite coding regions in *S. ricini* (Fig 4) which is on par with the previous study reported by Behura and Severson in 2012 in 25 different Insect species which includes *Bombyx mori*. In current study Arginine is the second most frequent amino acid in class II followed by Glycine. Similar results were reported by (Katti *et al.*, 2001), where Alanine and Glycine were found to be more and tyrosine is the least present at the coding regions having SSR motifs in *Drosophila melanogaster.*

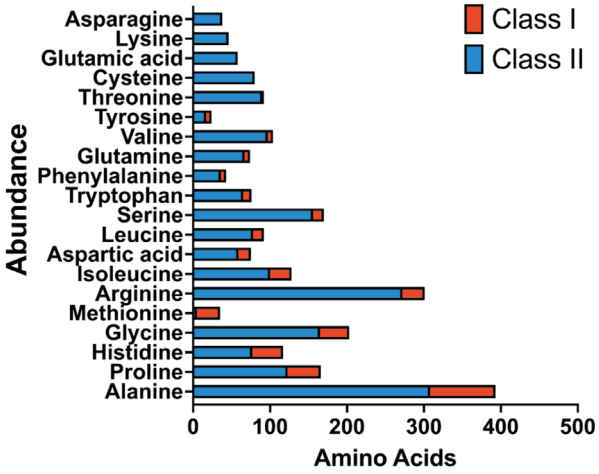

Fig. 4. The amino acid abundance in SSR loci in *Samia ricini* in both class 1 and class 2.

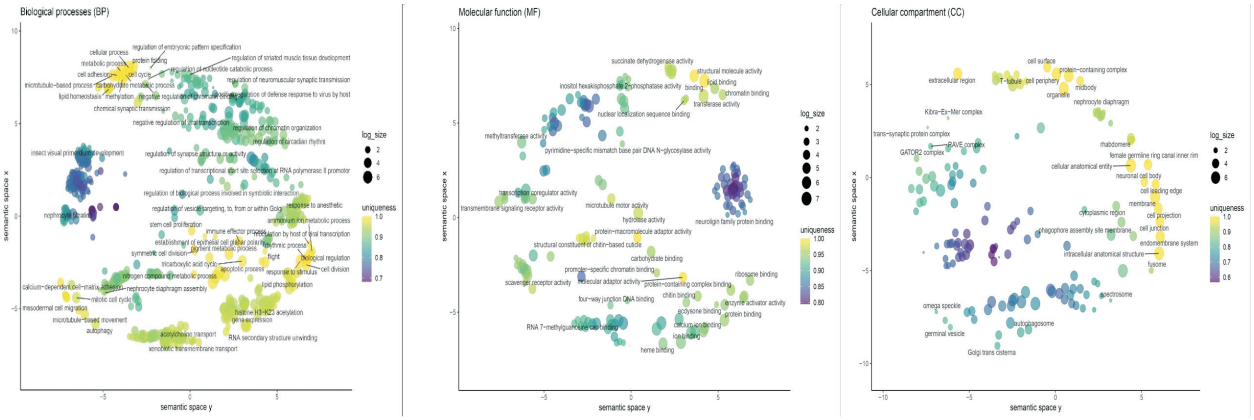

Fig. 3. Gene ontology scatterplot constructed by REVIGO depicting the cluster representation of the GO terms of given genes containing SSR motifs involved in A. Biological processes (BP), B. Molecular function (MF) and C. Cellular component (CC). Individual circle/bubble indicates cluster representatives, its colour indicates the p-value in right-hand corner; log size indicates the frequency of the GO term in the underlying GOA database. More functionally similar GO terms were closer in the scatterplot, but the semantic space units have no intrinsic meaning.

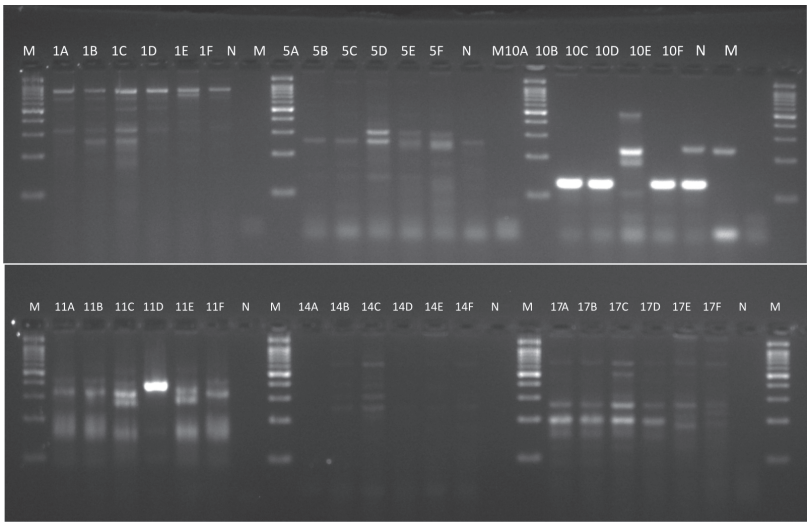

Fig. 5. The Eri silkworm *Samia ricini* has got Six morphotypes (A: Yellow Plain, B: Yellow Spotted, C: Yellow Zebra, D: Greenish Blue Plain, E: Greenish Blue Spotted, F: Greenish Blue Zebra), and the six SSR primers (1,5,10,11, 14 & 17) out of 20 Randomly selected primers gave the polymorphic band. M=Marker, N=non-template control.

Validation of the SSR primers

Over all 25,252 primer sets with a single amplicon (Table S2) were generated by using the Primer 3.0 software in *S. ricini* for 35,622 SSR containing sequences (SSR ≤18bp). Out of which 20 primer sets were randomly selected and synthesized for the validation (S5). All 20 primers were successfully amplified, with only six primers (1, 5, 10, 11, and 14) exhibiting polymorphic bands (Fig. 5) with 30% polymorphic percentage. The number of the bands given by a SSR primer sets ranging from 2-5.

Conclusion

The application of rapid detection protocols, like the polymerase chain reaction, along with the beneficial features of microsatellites, has resulted in their widespread use in agriculture. Microsatellites, which are relatively small in size but demonstrate high levels of polymorphism, contribute to the success of these protocols. These microsatellites are employed for various purposes including the characterization of genetic material, plant selection, and the construction of linkage maps for economically important traits. Given its emergence as a genetic model within lepidopteran insects, it is imperative to establish comprehensive data on genetic markers for further exploration in the Eri silkworm. Our study provides insights into the abundance of SSRsand their potential for cross-amplification across other Samia species. These SSR primers offer utility in population genetics, enabling the study of genetic relatedness among different species within the Genus *Samia*. Additionally, our gene ontology analysis contributes to the understanding of gene function, shedding light on the roles of various genes in the Eri silkworm. Considering the lack of promising breeds or hybrids for enhancing silk production in India's sericulture industry, the data generated from our study can be effectively leveraged for the improvement and development of Eri silkworm breeds and hybrids. Through these efforts, we aim to contribute to the advancement and sustainability of sericulture practices in India and beyond.

Acknowledgement

Authors thank the Central Silk Board, Ministry of textiles, Govt. of India, Bangalore for providing necessary infrastructure.

Statements and Declarations

Competing Interests: Authors does not have any financial interest related to the work submitted for publication. Authors declare no financial interests associated with the submitted work for publication, prioritizing unbiased dissemination of research. This ensures integrity and transparency in the scientific discourse, fostering credibility and trust within the academic community.

References

Aamir, M., Singh, V.K. and Meena M. 2017. Structural and Functional Insights into WRKY3 and WRKY4 Transcription Factors to Unravel the WRKY–DNA (W-Box) Complex Interaction in Tomato (*Solanum lycopersicum* L.). A Computational Approach. *Front Plant Sci.* 8: 819. https://doi.org/10.3389/fpls.2017.00819

Andrews, S. 2010. Fast QC: a quality control tool for high throughput sequence data

Behura, S.K. and Severson, D.W. 2012. Genome-wide Comparative Analysis of Simple Sequence Coding Repeats among 25 Insect Species. *Gene.* 504: 226-232. https://doi.org/10.1016/j.gene.2012.05.020

Beier, S., Thiel, T. and Münch, T. 2017. MISA-web: a web server for microsatellite prediction. *Bioinformatics.* 33: 2583-2585. https://doi.org/10.1093/bioinformatics/btx198

Bolger, A.M., Lohse, M. and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30: 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Botlagunta, M., Padhi, B., Ghosh, S. and Kundu, S. 2006. Genetic variation in ecoraces of tropical tasar silkworm, Antheraea mylitta D. using RFLF technique. *Current Science.* 90: 100-103.

Buchfink, B. 2021. Diamond v2.0.7 benchmark. https://doi.org/10.6084/m9.figshare.c.5053112.

Chakrabarty, S., Saha, A.K., Manna, B. and Bindroo, B. 2012. Pebrine disease in eri silkworm: A microscopic view. *Indian Silk.* 2: 16-17.

Conesa, A., Gotz, S. and Garcia-Gomez, J.M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 21: 3674-3676. https://doi.org/10.1093/bioinformatics/bti610

Ferro, P., dell'Eva, R. and Pfeffer, U. 2001. Are there CAG repeat expansion-related disorders outside the central nervous system? *Brain Res. Bull.* 56: 259-264. https://doi.org/10.1016/s0361-9230(01)00663-3

Foster, J.T, Bull, R.L. and Keim, P.S. 2011. Chapter 19 - Ricin Forensics: Comparisons to Microbial Forensics. In: Budowle B, Schutzer SE, Breeze RG, *et al*. (eds) *Microbial Forensics (Second Edition). Academic Press, San Diego*, pp 315-326

Hill, C.B., Cassin, A. and Keeble-Gagnère G. 2016. De novo transcriptome assembly and analysis of differentially expressed genes of two barley genotypes reveal root-zone-specific responses to salt exposure. *Sci Rep.* 6: 31558. https://doi.org/10.1038/srep31558

Hosamani, M., Basagoudanavar, S.H. and Sreenivasa, B.P. 2015. Eri silkworm (*Samia ricini*), a non-mulberry host system for AcMNPV mediated expression of recombinant proteins. *J Biotechnol.* 216: 76-81. https://doi.org/10.1016/j.jbiotec.2015.10.006

Humann, J.L., Lee, T., Ficklin, S. and Main, D. 2019. Structural and Functional Annotation of Eukaryotic Genomes with Gen SAS. In: Kollmar M (ed) *Gene Prediction: Methods and Protocols*. Springer, New York, NY, pp 29-51.

Jones, P., Binns, D. and Chang, H.Y. 2014. Inter Pro Scan 5: genome-scale protein function classification. *Bioinformatics.* 30: 1236-1240. https://doi.org/10.1093/bioinformatics/btu031

Jurka, J. and Pethiyagoda, C. 1995. Simple repetitive DNA sequences from primates: compilation and analysis. *J Mol Evol.* 40: 120–126. https://doi.org/10.1007/BF00167107

Kalia, R., Rai, M. and Kalia, S. 2011. Microsatellite markers: An overview of the recent progress in plants. *Euphytica.* 177: 309-334. https://doi.org/10.1007/s10681-010-0286-9

Katti, M.V., Ranjekar, P.K. and Gupta, V.S. 2001. Differential Distribution of Simple Sequence Repeats in Eukaryotic Genome Sequences. *Molecular Biology and Evolution.* 18: 1161-1167. https://doi.org/10.1093/oxfordjournals.molbev.a003903

Kim, M.C., Kim, N.Y., Seo, Y.R. and Kim, Y. 2017. A subset of micro RNAs defining the side population of a human malignant mesothelioma cell line. *Oncotarget.* 8: 42847-42856. https://doi.org/10.18632/oncotarget.17086

Kumara, R.R. 2023. Breeding in host plants of eri silkworm for rearing suitability. *Mysore J. Agric. Sci*. 57(3): 24-43.

Lee, J., Nishiyama, T. and Shigenobu, S. 2021. The genome sequence of *Samia ricini*, a new model species of lepidopteran insect. *Mol Ecol Resour.* 21: 327-339. https://doi.org/10.1111/1755-0998.13259

Li, B., Wang, X. and Hou, C.X. 2013. Genetic analysis of quantitative trait loci for cocoon and silk production quantity in *Bombyx mori* (Lepidoptera: Bombycidae). *European Journal of Entomology.* 110: 205-213. https://doi.org/10.14411/eje.2013.031

Li, Y. C., Korol, A. B., Fahima, T. and Nevo, E. 2004. Microsatellites Within Genes: Structure, Function, and Evolution. *Molecular Biology and Evolution.* 21: 991–1007. https://doi.org/10.1093/molbev/msh073

Lu, D., Huang, Y. and Naumann, S. 2022a. Mitochondrial genomes of two wild silkmoths, Samia watsoni and *Samia wangi* (Lepidoptera: Saturniidae), and their phylogenetic implications. *Eur J Entomol.* 119:337-353. https://doi.org/10.14411/eje.2022.035

Lu, D., Huang, Y., Naumann, S., Kitching, I., Xu, Z., Sun, Y. and Wang, X. 2022. Mitogenomic phylogenetics of the wild silkmoth genus *Samia* Hübner, 1819 (Lepidoptera: Saturniidae). *Authorea Preprints.*

Madesis, P., Ganopoulos, I. and Tsaftaris, A. 2013. Microsatellites: Evolution and Contribution. In: Kantartzi SK Microsatellites: Methods and Proto-

cols. *Humana Press, Totowa,* NJ, pp 1–13

Maegawa, K., Shibata, T. and Yamaguchi, R. 2018. Overexpression of a virus-like particle influenza vaccine in Eri silkworm pupae, using Autographa californica nuclear polyhedrosis virus and host-range expansion. *Arch Virol.* 163: 2787-2797. https://doi.org/10.1007/s00705-018-3941-4

Mandal, L., Verma, S.K., Sasmal, S. and Katara, J. 2018. Potential applications of molecular markers in plant. *Curr. Trends Biomed. Eng. Biosci.* 12: 85-87.

Meglécz, E., Anderson, S.J. and Bourguet, D. 2007. Microsatellite flanking region similarities among different loci within insect species. *Insect Mol Biol.* 16:175-185. https://doi.org/10.1111/j.1365-2583.2006.00713.x

Metz, S., Cabrera, J.M. and Rueda, E. 2016. FullSSR: Microsatellite Finder and Primer Designer. *Adv Bioinformatics* 2016:6040124. https://doi.org/10.1155/2016/6040124

Miao, X.X., Xub, S.J., Li, M.H., Li, W., Huang, J.H., Dai, F.Y., Marino, S.W., Mills, D.R., Zeng, P., Mita, K. and Jia, S.H. 2005. Simple sequence repeat-based consensus linkage map of *Bombyx mori. Proceedings of the National Academy of Sciences.* 102(45): 16303-16308.

Miao, X.X., Li, W.H. and Li, M.W. 2008. Inheritance and linkage analysis of co-dominant SSR markers on the Z chromosome of the silkworm (*Bombyx mori* L.). *Genetics Research.* 90: 151-156. https://doi.org/10.1017/S0016672308009221

Osanai-Futahashi, M., Suetsugu, Y., Mita, K. and Fujiwara, H. 2008. Genome-wide screening and characterization of transposable elements and their distribution analysis in the silkworm, *Bombyx mori. Insect Biochemistry and Molecular Biology.* 38: 1046-1057. https://doi.org/10.1016/j.ibmb.2008.05.012

Pérez-Martínez, A.L., Eguiarte, L.E. and Mercer, K.L. 2022. Genetic diversity, gene flow, and differentiation among wild, semiwild, and landrace chile pepper (*Capsicum annuum*) populations in Oaxaca, Mexico. *Am J Bot.* 109 : 1157-1176. https://doi.org/10.1002/ajb2.16019

Prasad, M.D., Muthulakshmi, M. and Madhu, M. 2005. Survey and Analysis of Microsatellites in the Silkworm, *Bombyx mori*. *Genetics.* 169: 197-214. https://doi.org/10.1534/genetics.104.031005

Qi, W.H., Yan, C. and Li, W.J. 2016. Distinct patterns of simple sequence repeats and GC distribution in intragenic and intergenic regions of primate genomes. *Aging (Albany NY).* 8: 2635-2650. https://doi.org/10.18632/aging.101025

Quinlan, A.R. and Hall, I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 26: 841-842. https://doi.org/10.1093/bioinformatics/btq033

Ranade, S.S., Lin, Y.C., Van de Peer, Y. and García-Gil, M.R. 2015. Comparative *in-silico* analysis of SSRs in coding regions of high confidence predicted genes in Norway spruce (*Picea abies*) and Loblolly pine (*Pinus taeda*). BMC Genet. 16: 149. https://doi.org/10.1186/s12863-015-0304

Ravinder, T., Kaki, S.S., Kunduru, K.R., Kanjilal, S., Rao, S.K., Swain, S.K. and Prasad, B.N. 2016. Isolation and characterization of phospholipids of eri-pupal oil from pupae grown on castor and tapioca leaves. *International Journal of Current Research and Review.* 8(15): 27.

Reddy, K.D., Abraham, E.G. and Nagaraju, J. 1999. Microsatellites in the silkworm, *Bombyx mori*: Abundance, polymorphism, and strain characterization. *Genome*. 42: 1057-1065. https://doi.org/10.1139/g99-027.

Renuka, G. and Shamitha, G. 2016. Genetic variation in ecoraces of tropical tasar silkworm, *Antheraea mylitta* using SSR markers. *J Genet.* 95: 777-785. https://doi.org/10.1007/s12041-016-0712-4

Supek, F., Bošnjak, M., Škunca, N. and Šmuc, T. 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS One.* 6: e21800. https://doi.org/10.1371/journal.pone.0021800

Untergasser, A., Cutcutache, I. and Koressaar, T. 2012. Primer3- new capabilities and interfaces. *Nucleic Acids Research.* 40: e115-e115. https://doi.org/10.1093/nar/gks596

Van't Hof, A.E., Brakefield, P.M., Saccheri, I.J. and Zwaan, B.J. 2007. Evolutionary dynamics of multilocus microsatellite arrangements in the genome of the butterfly *Bicyclus anynana*, with implications for other Lepidoptera. *Heredity.* 98: 320-328. https://doi.org/10.1038/sj.hdy.6800944

Vieira, M.L.C., Santini, L., Diniz, A.L. and Munhoz, C. de F. 2016. Microsatellite markers: what they mean and why they are so useful. *Genet Mol Biol.* 39: 312-328. https://doi.org/10.1590/1678-4685-GMB-2016-0027

Vijayan, K., Anuradha, H.J. and Nair, C.V. 2006. Genetic diversity and differentiation among populations of the Indian eri silkworm, *Samia cynthia ricini,* revealed by ISSR markers. *Journal of Insect Science.* 6: 30. https://doi.org/10.1673/2006_06_30.1

Wang, X.Y., Li M.W. and Zhao, Y.P. 2010. Mapping of non-lepis wing gene nlw in silkworm (*Bombyx mori)* using SSR and STS markers. *Yi Chuan.* 32: 54-58. https://doi.org/10.3724/sp.j.1005.2010.00054

Wani, S., Bhat, M., Buhroo, Z., Ganai, M. and Majid, N., 2013. Role of molecular markers in silkworm improvement. *International Journal of Recent Scientific Research*. 4(5): 515-523.

Zhan, S., Huang, J. and Guo, Q. 2009. An integrated genetic linkage map for silkworms with three parental combinations and its application to the mapping of single genes and QTL. BMC *Genomics.* 10: 389. https://doi.org/10.1186/1471-2164-10-389.